

The effectiveness of the Incredible Years pre-school parenting programme in the United Kingdom: a pragmatic randomised controlled trial

Morpeth, Louise; Blower, Sarah; Tobin, Kate; Taylor, Rod S.; Bywater, Tracey Jane; Edwards, Rhiannon; Axford, Nick; Lehtonen, Minna; Jones, Carys; Berry, Vashti

Child Care in Practice

DOI:

[10.1080/13575279.2016.1264366](https://doi.org/10.1080/13575279.2016.1264366)

Published: 01/04/2017

Peer reviewed version

[Cyswllt i'r cyhoeddiad / Link to publication](#)

Dyfyniad o'r fersiwn a gyhoeddwyd / Citation for published version (APA):

Morpeth, L., Blower, S., Tobin, K., Taylor, R. S., Bywater, T. J., Edwards, R., Axford, N., Lehtonen, M., Jones, C., & Berry, V. (2017). The effectiveness of the Incredible Years pre-school parenting programme in the United Kingdom: a pragmatic randomised controlled trial. *Child Care in Practice*, 23(2), 141-161. <https://doi.org/10.1080/13575279.2016.1264366>

Hawliau Cyffredinol / General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

The effectiveness of the Incredible Years pre-school parenting programme in the UK: a pragmatic randomised controlled trial

Authors:

Morpeth, Louise ^a
Blower, Sarah ^d
Tobin, Kate ^a
Taylor, Rod S. ^b
Bywater, Tracey ^d
Edwards, Rhiannon Tudor ^c
Axford, Nick ^a
Lehtonen, Minna ^a
Jones, Carys ^c
Berry, Vashti ^{a b}

Affiliations

^a Dartington Social Research Unit
^b Institute of Health Research, University of Exeter Medical School
^c Centre for Health Economics and Medicines Evaluation, Bangor University
^d University of York

Corresponding author

Dr. Vashti Berry, Room 2.05a, South Cloisters, St Luke's Campus, University of Exeter Medical School, EX1 2LU, v.berry@exeter.ac.uk

Trial registration site and number

www.controlled-trials.com: ISRCTN 48762440

Funding

This study was funded by Birmingham City Council (BCC), as part of its Brighter Futures strategy for children's services.

Acknowledgements

The authors would like to acknowledge the hard work and dedication of Birmingham City Council's (BCC) 'Brighter Futures' team, including the facilitators who ran parent groups and collected fidelity data, and Judy Hutchings, who oversaw the implementation of Incredible Years and provided supervision to the group leaders. Michael Little was the project Chief Investigator. We thank the parents who took part in the trial, contributing valuable data on their children's well-being. The North Wales Trials Unit (NORTH) provided the randomisation service for this study. Pat Linck provided invaluable assistance to Rhiannon Tudor Edwards on the cost-effectiveness study. Vashti Berry's time is supported by NIHR CLAHRC South West Peninsula (PenCLAHRC) and Tracey Bywater and Sarah Blower are part of the Healthy Children Healthy Families Theme of the NIHR CLAHRC Yorkshire and Humber.

Interests declared: None

Ethical clearance: The Warren House Group Ethics Committee and the North Wales NREC (reference 10/WNo01/29).

Abstract

The prognosis for children with early-onset conduct disorder is poor. Conduct disorder also has a social cost for families and communities, and an economic cost for society through the increased use of health, education, social, legal and detention services. This study sought to evaluate the impact of a pre-school group-based parenting programme for parents of young children at risk of conduct disorder. The Incredible Years (IY) BASIC programme was delivered to parents of children considered to be at risk of developing a conduct disorder and evaluated by pragmatic randomised controlled trial. Allocation to the intervention or a waiting list control was conducted on a 2:1 ratio, stratified by child's age, sex and children's centre catchment. Participants were parents of 161 children (110 intervention, 51 control) aged between 36 and 59 months (mean age 44 months, 63% boys) and at risk of a conduct disorder defined by scoring over the clinical cut-off on the Strengths and Difficulties Questionnaire (SDQ). The primary outcome measure was the SDQ, with secondary outcomes observed on the Eyberg Child Behaviour Inventory (ECBI) and the Arnold & O'Leary parenting scale (APS). At follow-up (six months post-baseline), compared to control, parents and children in the intervention group had made significant improvements in parenting and problem behaviour. The intent-to-treat analysis showed a mean between group difference in favour of IY on the SDQ total difficulties score at follow-up of 2.23 (95% CI: 0.13 to 4.34, $p < 0.05$, effect size: 0.50). IY was also superior to control on the ECBI (13.48, 95% CI: 2.31 to 22.64, $p < 0.05$, effect size: 0.37), and on the APS (0.29, 95% CI: 0.08 to 0.50, $p < 0.01$, effect size: 0.43). This study confirms the effectiveness of IY in a public system delivered with fidelity by regular children's centre staff, supporting findings from a similar trial in Wales. These results support the wider roll-out of IY to similar children.

Introduction

This paper reports on an experimental evaluation of the Incredible Years (IY) parenting programme, delivered to the parents of three- and four-year old children at risk of developing a conduct disorder attending universal early years provision. The study was part of a broader initiative that used a portfolio of evidence-based programmes (EBPs) to improve outcomes for children in Birmingham, UK (Little *et al.*, 2012).¹

Conduct disorder

Conduct disorders meet the diagnostic criteria captured in the ICD-10 (World Health Organization, 1992) or DSM-V (American Psychiatric Association, 2013). They are characterised by age-inappropriate disruptive and anti-social behaviours, notably high rates of oppositionality, defiance and aggression. In the school years, diagnostic symptoms include the violation of classroom and adult authority, including lying and cheating, and in the adolescent years they include violations of the law or community authority (Scott, 2015). With no intervention, conduct disorders are likely to get worse. 'Behaviour problems' are lower level and range from mildly disruptive to severely destructive behaviours. Such difficulties have a conduct dimension, characterised by aggression and defiance, as well as an emotional dimension, marked by negative affect and deficits in peer relationships and prosocial behaviours. They are mainly displayed in the home and school, particularly among younger children, but can also result in anti-social activities in the community, especially among adolescents.

¹ The ages of children receiving the different programmes in the portfolio were different, meaning that children whose parents were in receipt of IY would not have received any of the other three interventions.

Extent of the problem

In the UK, a national survey based on multiple report sources and diagnostic interviews indicated that 10% of children aged 5-15 years have had a mental disorder and half of these presented with clinically significant conduct disorders (ONS, 1999). The findings indicated a much higher proportion of boys (7%) showing evidence of a conduct-related disorder compared to girls (3%). Another UK study, based on children living in a disadvantaged neighbourhood in London, found that nearly 20% had conduct disorders (Attride-Sterling *et al.*, 2000). Rates of parent-reported adolescent conduct problems in the UK increased substantially in the 25-year period from 1974 to 1999, although the most recent evidence indicates a small drop between 1999 and 2004 (Collishaw, 2012).

In Birmingham, the UK's second largest city, where the present evaluation was conducted, a survey involving 500 parents of children aged 0-6 and using the parent version of the Strengths and Difficulties Questionnaire (SDQ) found a significantly greater proportion of children aged 5-6 years ($n=80$) were at risk of a conduct disorder (19%) compared to 7% in Great Britain as a whole (Hobbs *et al.*, 2011). A parallel survey, involving over 10,000 young people aged 7-18, and using the SDQ self-report measure, found that a significantly greater number of children aged 11-15 in the city ($n=3,293$) were likely to meet a clinical diagnosis for conduct disorder (21%) compared with Great Britain as a whole (11%) (Hobbs *et al.*, 2011). Provision for such children in the city was considered by local commissioners and managers to be limited, with most children's centres offering universal parenting programmes but very little that was targeted.

Reasons to address conduct problems in young children

There are at least three reasons to address conduct problems in young children. First, the prognosis for children with early-onset conduct disorders is poor (Copeland *et al.*, 2015; Scott, 2015). Short-term, these children typically develop high levels of unhappiness and low self-esteem (Fanti, 2013; Stone *et al.*, 2015), display low levels of social competence (Barker *et al.*, 2010) and may have difficulty in forming and maintaining friendships (Trentacosta and Fine, 2010). Long-term, children displaying early-onset conduct problems are at heightened risk for drug abuse and depression in their adolescent and adult years (Klostermann, *et al.*, 2016; Pingault, *et al.*, 2013; Stringaris, *et al.*, 2014). In addition, the early onset of aggressive behaviour, at least for boys, is one of the best predictors of anti-social and criminal behaviour in adolescence and adulthood, including violent offending (Cleary & Nixon, 2012; Hodgins *et al.*, 2013; Miller *et al.*, 2010).

Second, conduct disorders become more difficult to address as children grow older (Allen, 2011; Frick, 2012). It has been estimated that success rates in reducing conduct disorder are in the region of 75% for children under the age of 10 compared to a 25% success rate for adolescents (Patterson *et al.*, 1993). There is also evidence that if children with aggressive behavioural problems are not addressed by age eight, their learning and behavioural problems become less responsive to intervention and their condition is more likely to become a chronic disorder (Webster-Stratton *et al.*, 2004; Shaw, 2013).

Third, aside from the social cost for families and communities, conduct disorder has an economic cost for society through the increased use of health, education, social, legal and detention services (Piquero, 2013; Shivram *et al.*, 2009). For example, in the UK it has been estimated that the average cost per family of anti-social behaviour by young children is £15,382 per year (Knapp *et al.*, 1999), rising to £70,019 per person by the age of 28 (Scott *et al.*, 2001).

How to address conduct problems

A successful approach to addressing a problem requires a good understanding of the problem. Conduct disorder is the outcome of a number of risk and protective factors that shape the child's behaviour and social competence, making it a challenging issue to address. The key risk factors fall into three key domains: (i) child characteristics; (ii) family characteristics; and (iii) school and peer influences (Hutchings & Gardner, 2012). These are conceptually separate but are likely to be experienced by the child as seamless and interact with one another to produce an outcome that is diagnosable as conduct disorder (Scott, 2015). The emergence of conduct disorder is likely to involve a child whose temperament is impulsive and hyperactive (Bornovalova, *et al.*, 2014; Murray & Farrington, 2010). Parents respond to such a temperament with harsh or inconsistent discipline, which, in addition to offering a negative role model for the child, impedes the development of prosocial and cognitive skills (Patterson *et al.*, 1982; Hoeve *et al.*, 2009; Belsky & de Haan, 2010). This negative cycle may be intensified by parental stress and family conflict, which further reduces opportunities to stimulate the child's social and cognitive development (Can & Ginsburg-Block, 2016; Semke *et al.*, 2010). These home experiences may be further compounded at school as teachers find it difficult to cope with challenging behaviours and

the child finds it hard to make friends with other children, drawing them to towards similarly anti-social peers during later childhood (Boivin, *et al.*, 2013; Smith *et al.*, 2014).

The Incredible Years programme

This understanding of conduct disorder informed the design of IY, which is based on the principle that behaviour is learned through social interaction (Patterson, 1982). Its application can be summarised thus: 'The aim is to increase positive behaviours through a variety of rewards, whilst reducing unwanted behaviours through response cost or other strategies, resulting in their disappearance' (Hutchings & Webster-Stratton, 2004: 340). In practice, this means that interactions with the child, involving significant others such as parents and teachers, are designed to support and reinforce the child's pro-social and co-operative behaviours while simultaneously discouraging disruptive and confrontational behaviours.

The IY series comprises programmes for children, teachers and parents which can be delivered independently or simultaneously. The parent training programme indirectly encourages the child to develop social competence and reduce aggressive behaviours; the child training programme addresses children's impulsivity and poor social skills; and the teacher training programme aims to change ineffective teacher responses to children's poor behaviour (Hutchings & Gardner, 2012).

This study

The pre-school, group-based IY BASIC parenting programme for parents of children aged three-to-six years was implemented and evaluated undertaken by Hutchings et al. (2007), where it was offered in community settings such as children's centres, community centres and church halls to parents of children aged three and four years whose behaviour was screened and rated as problematic by parents. The study sought to test whether the programme would improve children's behaviour and social relationships at home and with other children, and whether it would improve parenting competence. It was hypothesised that there would be positive outcomes on all measures.

Methods

Design

The study was a pragmatic randomised controlled trial, designed to inform children's services' policy decisions by providing evidence of the real-world effectiveness of the IY intervention. Allocation to the intervention or a waiting list control was on a 2:1 ratio, stratified by child's age, sex and children's centre catchment.

Participants and procedure

To be eligible for the current trial, children had to be aged between 36 and 59 months (three-four years of age) and be rated by their parent(s) as above the total difficulties clinical cut-off score on the SDQ. Children receiving medication, specifically for behavioural problems, and those with an existing clinical diagnosis of ADHD or autism spectrum disorder, were not included in the trial. Two independent research ethics committees granted ethical approval.

Referrals to the programme were accepted from different professional groups, including health visitors, teachers, psychologists and family support workers. Partner agencies were briefed about the programme and evaluation via events such as NHS staff meetings, health visitor training days and school cluster meetings. Children's centre family support workers also acted as local champions, helping to engage potential referrers in order to explain the eligibility criteria and the anticipated benefits of the programme. Open days at local children's centres, nurseries and schools were also held to make parents aware of the groups and self-referrals were accepted. A number of outreach events were also held in public spaces that families often frequent, such as supermarkets, doctors' surgeries and shopping centres. All referrals were checked for eligibility before a baseline interview was carried out. A more detailed account of the parent recruitment and engagement process for this trial can be found in Axford et al. (2012).

Sample size

A sample size calculation was conducted before the trial began. It estimated that a sample of 162 children was required to detect a medium effect size of 0.45 with a power of 81% and alpha of 5%. The calculations were based on previous studies of the IY BASIC parent programme and assumed a within-person correlation of 0.58 and baseline value adjusted mean difference between intervention and control group at follow-up of 2.3 on the SDQ (with SD = 6).

Randomisation and blinding

Parents read an information sheet and signed a consent form to acknowledge that they would be randomly allocated to one of the two conditions and that if randomly allocated to the control condition, they would wait for approximately six months before they could attend a group.

Researchers inserted eligible participants' information into a web-based programme to enable the trials unit (NORTH) to conduct the randomisation. In order to meet the needs of real-world implementation, children were randomised on a 2:1 ratio using a dynamic allocation method, stratified by age and sex, which allowed children to be randomised at the point of recruitment. The algorithm allows sequential randomisation of individuals without allowing predictability of allocation. No sequence is generated up front; each participant is randomised based on the participants already randomised, ensuring a balance of treatment and controls overall, within stratum and stratification variables simultaneously (Russell *et al.*, 2011).

The researchers responsible for gathering the outcome data from parents were blinded to the trial condition of the participants at baseline and follow-up interviews. Follow-up blinding may have been compromised if parents discussed details of their parenting group with data collectors, but parents were respectfully asked not to discuss group attendance to reduce risk of potential bias.

Intervention condition

The BASIC group IV parent programme was the intervention under trial. It is described extensively elsewhere (Webster-Stratton & Hancock, 1998; Webster-Stratton, 2011). Briefly, it comprises a mixture of presentations by facilitators, individual and small group activities and homework, and focuses on harsh and ineffective parenting skills, poor monitoring and low cognitive stimulation. It seeks to: increase parents' positive communication skills, such as the use of praise and positive feedback to children, and reduce the use of criticism and unnecessary commands; improve parents' limit-setting skills by replacing smacking and other negative physical behaviours with non-violent discipline techniques and by promoting positive strategies such as ignoring the child's behaviour, allowing for logical consequences, providing re-direction, and developing problem-solving and empathy skills; improve parents' problem-solving skills and anger-management; and increase family support networks.

IY is the type of parent training programme recommended by the National Institute for Care and Health Excellence as suitable treatment or indicated prevention for child conduct disorder (NICE, 2013) and is considered a 'model' programme by well-established standards of evidence (Blueprints, 2012; Webster-Stratton *et al.*, 2001). This is based on a number of large randomised controlled trial evaluations demonstrating the success of the parenting programme on children's conduct problems (Furlong *et al.*, 2012; Menting *et al.*, 2013). In Wales, Hutchings *et al.* (2007) found an effect size (ES) of 0.33, measured on the Strengths and Difficulties Questionnaire (SDQ; Goodman, 1997). In England, using referrals to standard child mental health services, Scott *et al.* (2001) found an effect size of 0.51 on the same measure. Similar results have been demonstrated in voluntary sector organisations (ES = 0.48, Gardner *et al.*, 2006), disadvantaged communities in Ireland (ES = 0.48, McGilloway *et al.*, 2012) and Seattle – the original test site (Webster-Stratton *et al.*, 2004).

In the current study, parents attended the 12-week programme comprising a two-hour group session once a week led by at least two facilitators. It was run with a maximum of 12 parents in a group (partners of the primary caretaker were also invited). The groups were delivered across nine children's centres in the city. In total, 12 parent groups were run for intervention families across the nine centres between January 2009 and March 2011. Trained facilitators (n = 19) delivered the groups in pairs. Facilitators received weekly supervision by recognised IY trainers.

Intervention fidelity

The groups were monitored for implementation fidelity. There was a detailed manual for the 12-session programme and facilitators completed weekly checklists of content covered and key learning. The group sessions were video-recorded each week and two from each group were randomly selected and observed for quality of delivery and adherence to programme process and content. Parent satisfaction and engagement was measured using the IY satisfaction questionnaire and weekly attendance register.

Control condition

Given this was a pragmatic trial, parents allocated to the control condition were free to access any other services on offer as usual but were not offered the IY programme until after their 6-month follow-up interview. Following this interview, each control parent was invited to participate in a parenting group. Further groups were run between September 2010 and September 2011 to serve parents from the control group.

Measures

Demographics

A family demographic questionnaire, completed by the researcher during the interview at baseline and follow-up, gathered basic demographic information about the child and the family's circumstances. The version was adapted from the one used in the Hutchings et al. (2007) trial.

Outcomes

Outcome data were collected from the parents during interviews in their home (or at an alternative suitable venue if requested by the parent). Researchers interviewed intervention and control parents at baseline and six months after baseline, using the same set of measures at both time points. All measures are validated and standardised and have been used extensively in similar trials.

There were two measures of child behaviour. The primary outcome measure was the parent-rated SDQ (versions 3-4 and 4-16 year-olds). It comprises five subscales (each with 5 items) assessing hyperactivity, conduct, emotional difficulties, peer relations and pro-social behaviour, respectively, over the past six months. There are three response options for each item (0 = 'not true', 1 = 'somewhat true', 2='certainly true'). For each of the subscales the score can range from 0 to 10; a higher score indicates more problems for all subscales apart from the prosocial subscale, for which a higher score indicates more prosocial behaviour. A global 'total difficulties' score is calculated by summing the first four sub-scales (i.e. all except pro-social behaviour), with scores ranging from 0 to 40 (higher scores indicate greater problems). In the current study, the mean Cronbach's alpha on the global score was 0.68. There are cut-offs on all sub-scales and the total difficulties score to indicate a likely clinical disorder (Goodman et al., 2000a, 2000b), with the cut-off on the total difficulties score (≥ 17) being used as an eligibility criterion for the present study. All children were screened at baseline on the 3-4 year old SDQ measure. At post-test, the SDQ 3-4 or SDQ 4-16 version was used as appropriate, depending on age of the child.

An impact scale indicates the extent of the burden that the child's problem behaviour has on the family. This starts with a single question about whether the child has

difficulties with emotions, concentration, behaviour, or being able to get on with other people ('No', 'Yes – minor difficulties', 'Yes – definite difficulties', and 'Yes – severe difficulties'). If the answer is 'Yes', there are four additional questions, focusing respectively (in the parent version) on: chronicity, or duration ('less than a month', '1-5 months', '6-12 months', 'over a year'); distress to the child ('not at all', 'only a little', 'quite a lot', 'a great deal'); interference with the child's everyday life in terms of home life, friendships, classroom learning and leisure activities respectively ('not at all', 'only a little', 'quite a lot', 'a great deal'); and burden to the parent or family as a whole ('not at all', 'only a little', 'quite a lot', 'a great deal'). The parent report impact score is calculated by summing responses to overall distress and interference, generating a total score ranging from 0 to 10, where higher scores indicate greater impact.

The second measure of child behaviour, the Eyberg Child Behaviour Inventory (ECBI) (Eyberg & Ross, 1978), a 36-item parent-rated scale, was used as a secondary measure to assess the frequency and burden of children's conduct problems. Scores are calculated for the intensity of problem behaviours and the extent to which they are a problem. For intensity, parents indicate the frequency of 36 behaviours on a 7-point scale (*Never to Always*). The possible range of scores is 36 to 252. The problem subscale assesses whether parents consider the child's behaviour as a problem for themselves (*Yes/No*). The range of scores for this subscale is 0-36. In the current study, the mean Cronbach's alpha was 0.94 for the intensity scale and 0.93 for the problem scale, mirroring those found in an earlier study (Burns and Patterson, 2001).

Parenting competence was measured using the Arnold and O’Leary Parenting Scale (Arnold *et al.*, 1993), a 30-item measure of dysfunctional discipline practices. A score is summed for each of three sub-scales, namely laxness, over-reactivity and verbosity, as well as a total score. In the current study, the mean Cronbach’s alpha for the total score was 0.82.

Service use

In addition to the measures outlined above, the Client Service Receipt Inventory (CSRI) (Beecham & Knapp, 1992; Ridyard & Hughes, 2010) was administered to parents at each time point and completed by the researcher. It measures frequency of health, social, and educational service use and is used to estimate the associated costs of service use. Service use costs are reported in a separate paper (Edwards *et al.*, 2016).

Fidelity

The groups were also monitored for implementation fidelity, with a specific focus on dose, adherence and quality of delivery. In line with previous evaluations of IY, a combination of methods, including self-rated checklists and independent observation, were adopted. Dose was measured using a weekly attendance register (recommended dose ≥ 7 sessions).

Adherence was measured using a weekly self-rated checklist capturing the extent to which the session content and key learning, as documented in the detailed programme manual, were delivered by the group leader (Webster-Stratton, 1989). Examples of items on the checklist include whether the appropriate video vignettes are used, whether role-play activities are delivered and whether homework from the previous week has been reviewed. The checklists are scored in order to calculate a percentage score of adherence for each

session. A mean average of scores across the 12 weeks of the programme is produced to indicate overall level of adherence achieved by each group leader.

Adherence was also assessed by independent observation. All of the programme sessions were video-recorded and two recordings from each group were randomly selected for members of the research team to code for adherence to programme process and content as well as quality of delivery via the Parent Programme Implementation Checklist (Bywater, 2010). Levels of adherence are determined by coding for similar items found in the self-rated leader checklist.

With regards to quality of delivery, researchers code for a variety of techniques and methods that leaders are trained to employ in sessions, such as open-ended questioning and the inclusion of all parents in group discussion. The researchers also looked for evidence indicating the leader's enthusiasm and preparedness.

Data analysis

The primary analysis of all primary and secondary outcomes was conducted according to the principle of intent-to-treat and included all 161 participating families in the sample regardless of their IY programme attendance. Fourteen children were lost to follow-up (10 intervention and four control). Our preliminary analyses showed that the difference between the rate of attrition in the intervention and control groups was not statistically significant, and that the baseline primary outcome score between these families and those that responded was also not significant ($t = 1.73$, $p = 0.09$). Given this, data were imputed for these 14 cases using a Last Observation Carried Forward (LOCF) method, which assumes no change between baseline and follow-up for these children.

The primary analysis used an analysis of covariance (ANCOVA) approach to estimate post-intervention differences between the groups on parent-reported child outcomes, adjusting for area, treatment condition and baseline response values, and the stratification variables of age and sex. Goodness of statistical model fit was assessed by checking normality of residuals and absence of heteroscedasticity. Where statistically significant ($p < 0.05$) between group differences were seen, standardised effect sizes were calculated with Cohen's guidelines, using the standardised mean difference divided by the pooled standard deviation at post-test (Cohen, 1988). A secondary per protocol analysis was undertaken, involving the parents in the intervention group who completed the programme. Outcome results are reported as mean between group differences and 95% confidence intervals. All analyses were conducted using SPSS Version 20 for Mac.

Results

Baseline characteristics

A CONSORT diagram in Appendix A depicts the flow of referral, recruitment and retention to the trial, in line with recommended best practice for reporting RCTs (Schulz et al., 2010).

The final sample comprised 161 index children (intervention $n = 110$, control $n = 51$). There were 101 males and 60 females, with a mean age of 44 months ($SD = 6$) at baseline. Eight additional children were excluded from the final analyses as they violated protocol conditions. Statistical analyses revealed no significant differences on outcomes at baseline or follow-up between these eight violation cases and the sample as a whole.

The demographics for the sample as a whole and the treatment conditions are presented in Table 1, including those lost to follow-up. There was no evidence of a difference in demographics between the IY and control groups. The sample comprised a high proportion of low-income families: 50% of the families in the sample relied on benefits as their main source of income. All children, by the eligibility definition, met the clinical threshold on the SDQ total difficulties score but, not unusually, not all children demonstrated clinical levels on the sub-scales of conduct and emotional difficulties.

INSERT TABLE 1 ABOUT HERE

Fidelity

Overall, there were high levels of implementation fidelity across the nine participating children's centres. On average, group leaders and independent raters reported comparably high adherence scores (mean adherence 85% and 86% respectively). However, there was

greater variation in the levels of adherence between leaders when observed by independent raters (ranging between 78% and 89%). Independent ratings of the quality of programme delivery reached similarly high levels (mean average of 78%), with three IY leaders achieving 100% on this dimension of fidelity. Lower levels of fidelity were observed in relation to dose. Just over half of all the parents (51%) attended seven or more of the 12 sessions, though on average parents attended just six of the 12 sessions. Over a third of parents (38%) attended only one session or none at all.

The developers of IY incorporated strict fidelity controls in the training and accreditation process (Webster-Stratton, 2006). Programme leaders are required to participate in weekly supervision sessions and video recordings of every group they deliver are coded for their adherence and quality by the programme developers, sufficient levels of both lead to formal accreditation. These controls have likely contributed to the high levels of adherence and quality across the different facilitators. However, existing fidelity controls are not focused on parent attendance and the wraparound support that is often needed to help parents get to groups, such as the provision of transport and crèche facilities, which may have contributed to variation and lower-than-expected levels of parent attendance at IY groups.

There is currently no evidence of the minimum level of fidelity required in order to replicate the results of previous studies documenting the positive impact of the IY programme on child outcomes. It has been reported that higher levels of fidelity are associated with greater improvements in emotion and behaviour outcomes (Eames *et al.*, 2009; Mihalic, 2004). Thus, the relatively high levels of fidelity obtained in this study may have contributed to the significant improvement in child outcomes observed. The relationship between

fidelity and outcomes observed in this study is discussed in a separate paper (Blower & Berry, 2013).

Outcomes

The intention-to-treat analysis for the primary child outcome measure (SDQ) was run on all children, including imputed data for missing cases. Table 2 provides the mean scores at baseline and six months for the two conditions, as well as standardised mean differences, with 95% CI, and a converted Cohen's d effect size for the treatment where significant. It indicates that there was a significant difference between the intervention and control condition at follow-up on SDQ behavioural difficulties ($d=0.39$) but not emotional problems.

There were also significant differences between the intervention and control group at follow-up on the secondary outcome measure (ECBI). Intervention children showed reduced intensity of problems compared to the control group ($d=0.37$). There were also significant effects of the intervention on the measure of parenting competence at follow-up ($d= 0.36$ to 0.47). Mean levels of laxness, over-reactivity and verbosity are presented below for intervention and control groups.

INSERT TABLE 2 ABOUT HERE

INSERT TABLE 3 ABOUT HERE

Area-level analyses reveal no significant differences in children's outcomes by the area/group that they attended. Intervention families in the 12 different groups showed similar gains over their respective controls.

Per protocol analyses

The model was fitted for only those families that completed the intervention ($n = 55$); in other words, using data from parents who attended seven or more of the 12 sessions as the intervention group (Table 4). The difference between the control and intervention group was significantly greater once 'non-completers' were removed, with a larger reduction in global difficulties ($d=0.57$), conduct problems ($d=0.51$) and peer relationships ($d=0.54$).

INSERT TABLE 4 ABOUT HERE

However, this pattern is not due to parents of children with fewer difficulties at the start dropping out. A t-test comparison of the 'completer' and 'non-completer' intervention families indicates that there is not a significant difference between the children with 'completer' or 'non-completer' parents at baseline on conduct ($p = 0.64$) or emotion scores ($p = 0.78$).

Discussion

The evaluation results are promising, particularly given the complex nature and causes of conduct disorder. On a measure that detects changes in parent practices, there were greater reductions in reported negative parenting behaviours among the parents in the IY condition relative to the control. On a second set of measures that detect changes in child behaviour, there were larger reductions in behaviour problems and improvements in children's peer and family relationships in the IY condition relative to the control. These positive results may be a consequence of the quality of training and supervision arrangements for IY facilitators, which were delivered by the same team as in Wales. The results for the adherence and quality dimensions of fidelity were relatively high, although the dose was disappointing by contrast. A parallel analysis found that IY had a high probability of being cost-effective, shifting an additional 23% of children from above to below the SDQ clinical cut-off compared to the control condition, at a cost ranging from £1612 to £2418 per child, depending on the number of children per group (Edwards *et al.*, 2016).

These results largely replicate the findings of the Hutchings *et al.* (2007) study in Wales, which reported a strong impact on behavioural problems, measured by both the SDQ and ECBI. (Levels of take-up there were higher, with mean attendance greater than nine sessions compared with a mean of six in this study.) Moreover, these improvements were sustained at a longer-term follow-up (Bywater *et al.*, 2009) and found to be cost-effective for the clinical gains made (Edwards *et al.*, 2007).

Implications for practice

The consistency of effect sizes found in studies of this version of IY from the US, Wales, Ireland and now in England indicate that this is an early intervention programme that transports well across cultural contexts. It is also noteworthy that in the present study, the majority of children were from a minority ethnic group, so the intervention was delivered to parents from a range of backgrounds. The results on effectiveness and cost-effectiveness indicate that there is a strong case for implementing the parent version of IY more widely, a point supported by a recent meta-analysis of the programme (Menting et al., 2013). This is not to say that it should be treated as a panacea for conduct problems; as with any EBP addressing a public health issue, it needs to be complemented by a range of other practice and policy changes (Axford and Morpeth, 2013).

The challenges encountered with recruitment and retention in this study also need to be addressed. The screening required for the evaluation demonstrated that while children's centres reach thousands of disadvantaged families, those with children with significant impairments to development were under-represented. Where epidemiological data indicated that around 19% of children in the city present with behavioural problems at any given point (Hobbs *et al.*, 2011), less than 3% of the children's centre clients met that threshold (Axford *et al.*, 2012). There is no incentive for parents to put their difficult children on show in children's centres, and there are few incentives for managers of children's centres to reach out to the families in greatest need. In order to compensate for this, methods were introduced to increase the likelihood of parents struggling with children with behaviour difficulties using children's centres (Axford *et al.*, 2012). These methods should be a necessary part of any attempt to translate these results into mainstream practice, and arguably should be applied regardless of decisions to introduce IY into children's centres.

Implications for research

Three issues that emerged during this study demand further scrutiny. The first concerns transportability. This evaluation was part of a broader experiment to improve child outcomes. Of four evaluations of imported programmes in the city, IY was the only one to produce positive results.² This speaks to the difficulty of transporting the results of experiments across international boundaries and into mainstream children's services systems (Little, 2010). More research is therefore needed into whether there is a systematic bias in impact favouring home-grown over imported programmes, and what factors predict better transportability. In the case of parenting interventions, a recent meta-regression study found no difference between imported and home-grown programmes in effectiveness in reducing disruptive child behaviour, suggesting that interventions should be selected based on their evidence base and not on cultural specificity (Leijten et al., 2016).

The second issue is scale. The sponsor of this trial, the City Council, committed to scaling the delivery of the IY parenting programme across the city's children's centres. If successful, they will be the first public system in the world to scale an EBP. Conservative estimates, derived from epidemiological studies undertaken as part of the wider reform programme, suggest that 19% of all three- and four-year-olds in the city display behaviours that meet the IY entry criteria; this means that in the region of 4,800 children in the city could benefit from the intervention. Full implementation would help to reduce conduct disorder rates for three- and four-year-olds and meet national targets to reach more disadvantaged families.

² The others are reported elsewhere (Berry et al., 2015; Blower et al., 2016; Robling et al., 2016).

But in Wales, where dissemination of the IV series is at its most advanced, it is still not scaled to reach all who could benefit. Replicating good results across cultural contexts is an important advance in prevention and implementation science. Scaling those results to a broader population of children and families represents another frontier.

The third issue is fidelity. In this study, adherence and quality were relatively good but the dose was lower than recommended. The extent to which this matters, and whether and how different dimensions of fidelity need to be weighted, deserves further exploration (see Axford et al., 2017).

Strengths and limitations

This study has several methodological strengths, including adequate statistical power, the use of an independent trials unit to oversee randomisation, and the use of both intention-to-treat and per protocol analyses to provide an accurate estimate of the clinical effectiveness of the programme in a community-based setting (McGilloway *et al.*, 2012; Bywater *et al.*, 2009). Furthermore, a number of different measures were adopted to monitor implementation fidelity, including a combination of self-report checklists and independent observations, and the results indicate that the intervention was delivered with high levels of both adherence and quality.

The study also has limitations. Despite efforts to minimise attrition, 14 participants were lost to follow-up. There appear to be a larger proportion of girls and children meeting the clinical cut-off for behaviour disorders on the SDQ in the group lost to follow-up than the sample as a whole, though the differences are marginal. Another possible limitation of the

study is the relatively short follow-up (six months after baseline), although this length of time has been typical in studies of this nature (Hutchings *et al.*, 2007). Further, this study reported relatively low levels of dose compared to other evaluations of IY, such as the 85% attendance reported in Hutchings *et al.* (2007) and 88% in Webster-Stratton (1998). However, similar levels of parent attendance were reported in McGilloway *et al.* (2012) and are likely to be representative of ‘real-world’ attendance. Finally, the study relies on parent-report outcome data. Parents were unaware of their group allocation at baseline, but due to the nature of the intervention they were aware of whether they were in the intervention or control condition at follow-up. This may have biased their responses. With additional resource and time, a set of direct observation measures could have been used to complement the parent report data. That said, previous evaluations of IY have adopted mixed methods and found parent self-report data on outcomes to be consistent with independently observed data (e.g. Hutchings *et al.*, 2007 and McGilloway *et al.*, 2012). The effects reported in this study are comparable in magnitude to those found in studies that have used mixed methods.

References

- Allen, G. (2011). *Early Intervention: The Next Steps*, London: HM Government.
- American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.). Arlington, VA: American Psychiatric Publishing.
- Arnold, D. S., O’Leary, S. G., Wolff, L. S. & Acker, M. M. (1993). The parenting scale: a measure of dysfunctional parenting in discipline situations. *Psychological Assessment*, 5, 137-44.

Attride-Sterling, J., Davis, H., Day, C., & Sclare, I. (2000). *An Assessment of the Psychosocial Needs of Children and Families in Lewisham: Final Report*. London: South London and Maudsley NHS Trust.

Axford, N., Lehtonen, M., Tobin, K., Kaoukji, D. & Berry, V. (2012). Engaging parents in parenting programs: lessons from research and practice. *Children and Youth Services Review*, 34(10), 2061-2071.

Axford, N. and Morpeth, L. (2013) 'Evidence-based programs in children's services: a critical appraisal', *Children and Youth Services Review* 35(1), 268-277.

Axford, N., Bywater, T., Blower, S., Berry, V., Baker, V. & Morpeth, L. (2017) Critical factors in the implementation of parenting programmes: fidelity, adaptation and promoting quality. In Dixon, L., Perkins, D., Craig, L. A. & Hamilton-Giachritsis, C. (Eds) *What Works in Child Protection: An Evidence-Based Approach to Assessment and Intervention in Care Proceedings*. Wiley-Blackwell.

Barker, E. D., Oliver, B. R., & Maughan, B. (2010). Co-occurring problems of early onset persistent, childhood limited, and adolescent onset conduct problem youth. *Journal of Child Psychology and Psychiatry*, 51(11), 1217-1226.

Beecham, J., & Knapp, M. (1992). Costing psychiatric interventions. In G. Thornicroft, C. Brewin & J. Wings (Eds.) *Measuring mental health needs* (pp. 179-190). Oxford: Oxford University Press.

Belsky, J. & de Haan, M. (2010). Annual research review: parenting and children's brain development – the end of the beginning. *Journal of Child Psychology and Psychiatry*, 52, 409-428.

Berry, V., Axford, N., Blower, S., Taylor, R. S., Edwards, R. T., Tobin, K., Jones, C. & Bywater, T. (2016). The effectiveness and micro-costing analysis of a universal, school-based, social-emotional learning programme in the UK: a cluster-randomised controlled trial. *School Mental Health* 8 (2), 238-256.

Blower, S. & Berry, V. (2013). *Fidelity as a moderator of programme success: strengthening influencing factors*. Paris: European Society for Prevention Research.

Blower, S. et al. (2016). The effectiveness and micro-costing analysis of the Triple-P parenting programme: An independent UK randomised controlled trial

Blueprints (2012). Blueprints for Violence Prevention Website: The Incredible Years Series.

Retrieved October 16, 2012, from

<http://www.colorado.edu/cspv/blueprints/modelprograms/IYS.html>

Boivin, M., Brendgen, M., Vitaro, F., Forget-Dubois, N., Feng, B., Tremblay, R. E., & Dionne, G. (2013). Evidence of gene–environment correlation for peer difficulties: Disruptive behaviors predict early peer relation difficulties in school through genetic effects. *Development and psychopathology*, 25(01), 79-92.

Bornovalova, M. A., Cummings, J. R., Hunt, E., Blazei, R., Malone, S., & Iacono, W. G. (2014). Understanding the relative contributions of direct environmental effects and passive genotype-environment correlations in the association between familial risk factors and child disruptive behavior disorders. *Psychological medicine*, 44(4), 831-844.

Bywater, T. (2010). The Parent Programme Implementation Checklist: A training manual. Unpublished.

Bywater, T., Hutchings, J., Daley, D., Whitaker, C., Tien Yeo, S., Jones, K., Eames, C. & Edwards, R.T. (2009). Long-term effectiveness of a parenting intervention for children at risk of developing conduct disorder. *The British Journal of Psychiatry*, 195, 318-324.

Can, D. D., & Ginsburg-Block, M. (2016). Parenting stress and home-based literacy interactions in low-income preschool families. *Journal of Applied Developmental Psychology*, 46, 51-62.

Cleary, A., & Nixon, E. (2012). Early adult outcomes for Irish children with behavioural difficulties. *International Journal of Social Psychiatry*, 58(6), 643-651.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale NJ: Lawrence Erlbaum Associates.

Collishaw, S. (2012) Time trends in young people's emotional and behavioural problems, 1975-2005. In A. Hagell (Eds.) *Changing Adolescence: Social Trends and Mental Health*. Bristol: Policy Press.

Copeland, W. E., Wolke, D., Shanahan, L., & Costello, E. J. (2015). Adult functional outcomes of common childhood psychiatric problems: a prospective, longitudinal study. *JAMA Psychiatry*, 72(9), 892-899.

Eames, C., Daley, D., Hutchings, J., Whitaker, C. J., Jones, K., Hughes, J. C. & Bywater, T. (2009). Treatment fidelity as a predictor of behaviour change in parents attending group-based parent training. *Child: Care, Health and Development*, 35(5), 603-612.

Edwards, R. T., Ó Céilleachair, A., Bywater, T., Hughes, D. A. & Hutchings, J. (2007). Parenting programme for parents of children at risk of developing conduct disorder: cost effectiveness analysis. *British Medical Journal*, 10, 334-682.

Edwards, R. T., Jones, C., Berry, V., Charles, J., Linck, P., Bywater, T. & Hutchings, J. (2016) Incredible Years parenting programmes: cost-effectiveness and implementation. *Journal of Children's Services*, 11(1), 54-72.

Eyberg, S. M. & Ross, A. (1978). Assessment of child behavior problems: The validation of a new inventory. *Journal of Clinical Child Psychology*, 7(2), 113-116.

Fanti, K. A. (2013). Individual, social, and behavioral factors associated with co-occurring conduct problems and callous-unemotional traits. *Journal of abnormal child psychology*, 41(5), 811-824.

Frick, P. J. (2012). Developmental pathways to conduct disorder: Implications for future directions in research, assessment, and treatment. *Journal of clinical child & adolescent psychology*, 41(3), 378-389.

Furlong, M., McGilloway, S., Bywater, T., Hutchings, J., Smith, S. M. & Donnelly, M. (2012). Behavioural and cognitive-behavioural group-based parenting programmes for early-onset conduct problems in children aged 3 to 12 years. *Cochrane Database of Systematic Reviews*, 2.

Gardner, F., Burton, J. & Klimes, I. (2006). Randomised controlled trial of a parenting intervention in the voluntary sector for reducing child conduct problems: outcomes and mechanisms of change. *Journal of Child Psychology and Psychiatry*, 47, 1123-32.

Gardner, F. (1992). Parent-child interaction and conduct disorder, *Educational Psychology Review*, 4(2), 135-163.

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A Research Note. *Journal of Child Psychology and Psychiatry*, 38, 581-586.

Goodman, R., Ford, T., Simmons, H., Gatward, R. & Meltzer H. (2000b) Using the Strengths and Difficulties Questionnaire (SDQ) to screen for child psychiatric disorders in a community sample. *British Journal of Psychiatry*, 177, 534-539.

Goodman, R., Renfrew, D., & Mullick, M. (2000a) Predicting type of psychiatric disorder from Strengths and Difficulties Questionnaire (SDQ) scores in child mental health clinics in London and Dhaka. *European Child and Adolescent Psychiatry*, 9, 129-134.

Hartman, R. R., Stage, S. A., & Webster-Stratton, C. (2003). A Growth Curve Analysis of Parent Training Outcomes: Examining the Influence of Child Risk Factors (Inattention, Impulsivity, and Hyperactivity Problems), Parental and Family Risk Factors. *Journal of Child Psychology and Psychiatry*, 44(3), 388-398.

Hobbs, T., Axford, N. & Jodrell, D. (2011). Getting the measure of child health and development outcomes (2): the picture for a local authority in England. *Child Indicators Research*, 4(1), 81-100.

Hodgins, S., Larm, P., Ellenbogen, M., Vitaro, F., & Tremblay, R. E. (2013). Teachers' ratings of childhood behaviours predict adolescent and adult crime among 3016 males and females. *The Canadian Journal of Psychiatry*, 58(3), 143-150.

Hoeve, M., Dubas J. S. & Eichelsheim, V. I. (2009). The relationship between parenting and delinquency: a meta-analysis. *Journal of Abnormal Child Psychology*, 37, 749-775.

Hutchings, J. & Gardner, F. (2012). Support from the start: effective programmes for three to eight year-olds. *Journal of Children's Services*, 7(1), 29-40.

Hutchings, J. & Webster-Stratton, C. (2004). Community-based support for parents. In M. Hoghugi & N. Long (Eds.) *Handbook of Parenting: Theory and Research for Practice* (pp.334-351). London: Sage Publications.

Hutchings, J., Bywater, T., Daley, D., Gardner, F., Whitaker, C., Jones, K., Eames, C. & Edwards, R. T. (2007). Parenting Intervention in Sure Start Services for Children at Risk of Developing Conduct Disorder: Pragmatic Randomized Controlled Trial. *British Medical Journal*, 334, 1-7.

Klostermann, S., Connell, A., & Stormshak, E. A. (2016). Gender differences in the developmental links between conduct problems and depression across early adolescence. *Journal of Research on Adolescence*, 26, 76–89.

Knapp, M., Scott, S., & Davies, J. (1999). The cost of antisocial behaviour in young children. *Clinical Psychology and Psychiatry*, 4(4), 457-473.

Leijten, P., Melendez-Torres G. J., Knerr, W. & Gardner, F. (2016). Transported versus homegrown parenting interventions for reducing disruptive child behavior: a multilevel

meta-regression study. *Journal of the American Academy of Child & Adolescent Psychiatry*, doi:10.1016/j.jaac.2016.05.003.

Little, M. (2010). *Proof Positive*. London: Demos.

Little, M., Berry, V., Morpeth, L., Blower, S., Axford, N., Taylor, R., Bywater, T., Lehtonen, M. & Tobin, K. (2012). The impact of three evidence-based programmes delivered in public systems in Birmingham, UK. *International Journal of Conflict and Violence*, 6(2), 260-272.

McGilloway, S., Mhaille, G., Bywater, T., Furlong, M., Leckey, Y., Kelly, P. & Donnelly, M. (2012). A parenting intervention for childhood behavioural problems: A randomized controlled trial in disadvantaged community based settings. *Journal of Consulting and Clinical Psychology*, 80, 116-127.

Menting, A. T., Orobio de Castro, B. and Matthys, W. (2013) Effectiveness of the Incredible Years parent training to modify disruptive and prosocial child behavior: a meta-analytic review. *Clinical Psychology Review* 33(8), 901-913.

Mihalic S. (2004). The importance of implementation fidelity. *Emotional & Behavioral Disorders in Youth*, 4, 83–105.

Miller, S., Malone, P. S., Dodge, K. A., & Conduct Problems Prevention Research Group. (2010). Developmental trajectories of boys' and girls' delinquency: Sex differences and links to later adolescent outcomes. *Journal of abnormal child psychology*, 38(7), 1021-1032.

Murray, J., & Farrington, D. P. (2010). Risk factors for conduct disorder and delinquency: key findings from longitudinal studies. *The Canadian Journal of Psychiatry*, 55(10), 633-642.

NICE (National Institute for Health and care Excellence). (2013). *Antisocial Behaviour and Conduct Disorders in Children and Young People: Recognition and Management*. Clinical Guideline 158. London: NICE.

ONS (Office for National Statistics). (1999). *The mental health of children and adolescents in Great Britain: Summary Report*. London: Office for National Statistics.

Patterson, G. R., Chamberlain, P., & Reid, J. B. (1982). A comparative evaluation of parent training procedures. *Behavior Therapy*, 13, 638-650.

Patterson, G. R., Dishion, T. J., & Chamberlain, P. (1993). Outcomes and methodological issues relating to treatment of antisocial children. In T. R. Giles (Eds.), *Handbook of effective psychotherapy* (pp. 43-88). New York: Plenum.

Pingault, J. B., Côté, S. M., Galéra, C., Genolini, C., Falissard, B., Vitaro, F., & Tremblay, R. E. (2013). Childhood trajectories of inattention, hyperactivity and oppositional behaviors and prediction of substance abuse/dependence: a 15-year longitudinal population-based study. *Molecular Psychiatry*, 18(7), 806-812.

Piquero, A. R., Jennings, W. G., & Farrington, D. (2011). The monetary costs of crime to middle adulthood: Findings from the Cambridge study in delinquent development. *Journal of Research in Crime and Delinquency*, 50, 53-74.

Dyard, C. H., & Hughes, D. A. (2010). Methods for the collection of resource use data within clinical trials: a systematic review of studies funded by the UK Health Technology Assessment Program. *Value in Health*, 13, 867-872.

Robling, M., Bekkers, M-J., Bell, K. et al. (2016). Effectiveness of a nurse-led intensive home-visitation programme for first-time teenage mothers (Building Blocks): a pragmatic randomised controlled trial. *Lancet* 387, 146-155.

Russell, D., Hoare, Z. S., Whitaker, R., Whitaker, C. J. & Russell, I.T. (2011). Generalized method for adaptive randomization in clinical trials. *Statistics in Medicine*, 30 (9), 922-934.

Schulz, K.F., Altman, D.G. & Moher, D. for the CONSORT Group. (2010). CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *British Medical Journal*, 340:c332.

Scott, S. (2015). Oppositional and conduct disorders. In *Rutter's Child and Adolescent Psychiatry: Sixth Edition*. (pp. 911-930). John Wiley and Sons Ltd.

Scott, S., Knapp, M., Henderson, J. & Maughan, B. (2001). Financial cost of social exclusion: Follow-up study of antisocial children into adulthood. *British Medical Journal*, 323, 191-193.

Scott, S., Spender, Q., Doolan, M. Jacobs, B. & Aspland, H. (2001). Multicentre controlled trial of parenting groups for childhood antisocial behaviour in clinical practice. *British Medical Journal*, 323, 1-7.

Semke, C. A., Garbacz, S. A., Kwon, K., Sheridan, S. M., & Woods, K. E. (2010). Family involvement for children with disruptive behaviors: The role of parenting stress and motivational beliefs. *Journal of School Psychology*, 48(4), 293-312.

Shaw, D. S. (2013). Future directions for research on the development and prevention of early conduct problems. *Journal of Clinical Child & Adolescent Psychology*, 42(3), 418-428.

Shivram, R., Bankart, J., Meltzer, H., Ford, T., Vostanis, P., & Goodman, R. (2009). Service utilization by children with conduct disorders: findings from the 2004 Great Britain child mental health survey. *European child & adolescent psychiatry*, 18(9), 555-563.

Smith, J. D., Dishion, T. J., Shaw, D. S., Wilson, M. N., Winter, C. C., & Patterson, G. R. (2014). Coercive family process and early-onset conduct problems from age 2 to school entry. *Development and psychopathology*, 26(4 0 1), 917.

Stone, L. L., Otten, R., Engels, R. C., Kuijpers, R. C., & Janssens, J. M. (2015, October). Relations Between Internalizing and Externalizing Problems in Early Childhood. In *Child & Youth Care Forum* (Vol. 44, No. 5, pp. 635-653). Springer US.

Stringaris, A., Lewis, G., & Maughan, B. (2014). Developmental pathways from childhood conduct problems to early adult depression: findings from the ALSPAC cohort. *The British Journal of Psychiatry*, 205(1), 17-23.

Trentacosta, C. J., & Fine, S. E. (2010). Emotion knowledge, social competence, and behavior problems in childhood and adolescence: A meta-analytic review. *Social Development*, 19(1), 1-29.

Webster-Stratton C., Reid M. J. & Hammond M. (2004). Treating children with early-onset conduct problems: intervention outcomes for parent, child, and teacher training. *Journal of Clinical Child and Adolescent Psychology*, 33(1), 105–124.

Webster-Stratton, C. (1989). *The Incredible Years: The Parents and Children Series*. Seattle, WA, USA: IY.

Webster-Stratton, C. (1998). Preventing Conduct Problems in Head Start Children: Strengthening parenting Competencies. *Journal of Consulting and Clinical Psychology*, 66(5), 715-730.

Webster-Stratton, C. (2006). Treating Children with Early-onset Conduct Problems: Key Ingredients to Implementing the Incredible Years Programs with Fidelity. In K. T. Neill (Ed.), *Helping Others Help Children: Clinical Supervision of Child Psychotherapy*. New York: American Psychological Association.

Webster-Stratton, C. & Hancock, L. (1998). Training for parents of young children with conduct problems: Contents, methods and therapeutic processes. In G.E. Schaefer & J.M. Breisemeister (Eds.), *Handbook of Parent Training* (pp. 98-152). New York: John Wiley.

Webster-Stratton, C. (2011). *The Incredible Years: Parents, teachers, and children's training series: program content, methods, research and dissemination 1980-2011*. Seattle, WA: Incredible Years.

Webster-Stratton, C., Mihalic, S., Fagan, A., Arnold, D., Taylor, T., & Tingley, C. (2001). *The Incredible Years: Parent, Teacher And Child Training Series: Blueprints for Violence Prevention, Book Eleven*. Blueprints for Violence Prevention Series (D.S. Elliott, Series Editor). Boulder, CO: Center for the Study and Prevention of Violence, Institute of Behavioral Science, University of Colorado.

Webster-Stratton, C., Reid, M. J., & Hammond, M. (2004). Treating children with early-onset conduct problems: Intervention outcomes for parent, child, and teacher training. *Journal of Clinical Child and Adolescent Psychology*, 33(1), 105-124.

World Health Organization. (1992). *International Statistical Classification of Diseases and Related Health Problems, Tenth Revision. Vol. 1*. Geneva: World Health Organization.

Figure 1: Flow of Incredible Years participants through the trial and follow-up at 6 months after baseline

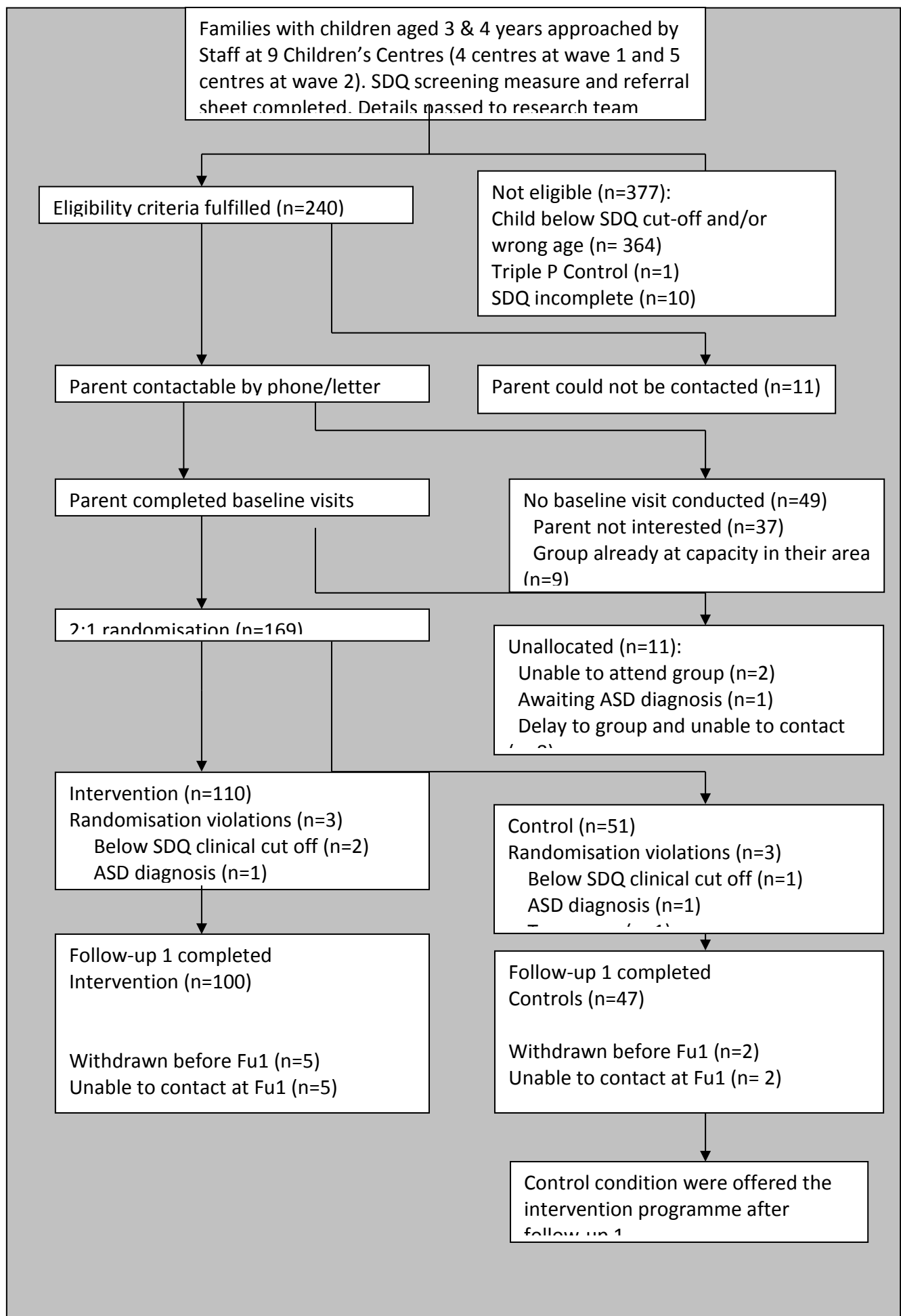


Table 1: Baseline demographics of the sample and clinical characteristics

	Waiting list controls (n = 47)	Intervention (n = 100)	Lost to follow-up	
			Controls (n = 4)	Intervention (n = 10)
No (%) boys	33 (70%)	63 (63%)	3 (75%)	2 (20%)
No (%) English second language	13 (28%)	29 (29%)	1 (25%)	3 (33%)
Mean age (months)	43	44	47	45
Ethnicity:				
No (%) White	16 (34%)	50 (50%)	1 (25%)	2 (20%)
No (%) Indian subcontinent	6 (13%)	14 (14%)	1 (25%)	1 (10%)
No (%) South East Asian	5 (11%)	10 (10%)	0	0
No (%) Mixed parentage	8 (17%)	6 (6%)	0	2 (20%)
No (%) Black African-Caribbean	1 (2%)	2 (2%)	1 (25%)	2 (20%)
No (%) Other	11 (23%)	18 (18%)	1 (25%)	3 (30%)
No (%) meeting clinical cut-off on behaviour*	44 (94%)	92 (92%)	2 (50%)	8 (80%)
No (%) meeting clinical cut-off on emotions*	23 (49%)	56 (56%)	3 (75%)	5 (50%)

* Cut-off = score of 4 or more on the SDQ behaviour and 5 or more on emotional difficulties

Table 2: ANCOVA results for intent-to-treat analysis on primary outcome

Mean (SD) raw scores						
Child Measure (cut-off score)	Control (n=51)		Intervention (n=110)		Est. mean diff. (95% CI) ³	Effect Size (d)
	Baseline	6 mth	Baseline	6 mth		
SDQ ⁴ conduct problems (≥4)	6.53 (2.1)	4.43 (2.7)	6.29 (2.0)	3.62 (2.1)	0.78* (0.05 to 1.51)	0.39
SDQ emotion problems (≥5)	4.85 (2.5)	3.61 (2.6)	4.79 (2.4)	3.30 (2.3)	0.36 (-0.36 to 1.07)	
SDQ hyperactivity (≥7)	7.66 (1.9)	6.18 (2.4)	7.67 (1.9)	5.83 (2.5)	0.40 (-0.36 to 1.17)	
SDQ peer problems (≥4)	4.47 (1.9)	3.39 (2.1)	4.23 (1.8)	2.69 (1.8)	0.71* (0.85 to 1.34)	0.39
SDQ pro-social behaviour (≤4)	5.19 (2.2)	6.35 (2.2)	5.72 (2.1)	6.77 (2.1)	-0.22 (-0.84 to 0.40)	
SDQ total difficulties (≥17)	23.50 (4.5)	17.60 (7.3)	22.98 (4.4)	15.44 (6.0)	2.23* (0.13 to 4.34)	0.50
SDQ impact (≥2)	0.92 (1.4)	0.58 (1.2)	0.59 (1.1)	0.14 (0.5)	0.37** (0.10 to 0.63)	0.31

* indicates significant at $p < .05$

** indicates significant at $p < .01$

³ Difference in mean follow-up scores between intervention and waiting list control conditions, measured by analysis of covariance adjusted for baseline score, age of child, sex and area.

⁴ SDQ Strengths and Difficulties Questionnaire (on all scales, higher scores equal greater problems, except for pro-social behaviour, which is measured positively).

Table 3: ANCOVA results for intent-to-treat analysis on secondary outcomes

Mean (SD) raw scores						
Child Measure (cut-off score)	Control (n=51)		Intervention (n=110)		Est. mean diff. (95% CI) ⁵	Effect Size (d)
	Baseline	6 mth	Baseline	6 mth		
ECBI-I (127) ⁶	143.86 (38.5)	134.35 (42.3)	142.70 (35.7)	123.10 (34.8)	13.48* (2.31 to 22.64)	0.37
ECBI-P (11) ⁷	17.31 (9.3)	14.33 (9.8)	16.71 (8.8)	11.24 (9.0)	2.62 (-0.07 to 5.32)	
APS total ⁸	3.58 (0.8)	3.32 (0.8)	3.49 (0.6)	3.01 (0.8)	0.29** (0.08 to 0.50)	0.43
APS laxness	3.79 (1.3)	3.43 (1.2)	3.58 (1.2)	3.04 (1.1)	0.30 (-0.01 to 0.61)	
APS verbosity	4.15 (0.9)	4.01 (1.0)	4.26 (0.9)	3.68 (1.0)	0.42** (0.12 to 0.72)	0.47
APS over-react	2.90 (1.0)	2.71 (1.1)	2.78 (0.8)	2.36 (0.8)	0.31* (0.06 to 0.57)	0.36

* indicates significant at $p < .05$

** indicates significant at $p < .01$

⁵ Difference in mean follow-up scores between intervention and waiting list control conditions, measured by analysis of covariance adjusted for baseline score, age of child, sex and area.

⁶ ECBI-I Eyberg Child Behaviour Inventory – Intensity Scale (higher scores equate to greater problems).

⁷ ECBI-P Eyberg Child Behaviour Inventory – Problem Scale (higher scores equate to greater problems).

⁸ APS Arnold and O’Leary Parenting Scale (and 3 sub-scales) - higher scores equate to greater problems.

Table 4: Per protocol analysis for child outcomes

Mean (SD) raw scores						
Child Measure (cut-off)	Control (n= 47)		Intervention Completers (n=55)		Est. mean diff. (95% CI)	Effect size (d)
	Baseline	6 mth	Baseline	6 mth		
SDQ conduct problems (4)	6.74 (2.0)	4.47 (2.8)	6.16 (1.9)	2.95 (1.7)	1.051* (0.19 to 1.91)	0.51
SDQ emotion problems (5)	4.77 (2.5)	3.43 (2.5)	5.18 (2.3)	2.93 (1.8)	0.666 (-0.15 to 1.48)	
SDQ hyperactivity (7)	7.76 (1.8)	6.15 (2.5)	7.79 (2.1)	5.47 (2.5)	0.525 (-0.36 to 1.41)	
SDQ peer problems (4)	4.40 (1.9)	3.23 (2.1)	4.57 (1.8)	2.36 (1.6)	0.956** (0.23 to 1.68)	0.54
SDQ pro-social behaviour (<4)	5.12 (2.3)	6.38 (2.3)	5.45 (2.3)	7.09 (2.1)	-0.462 (-1.19 to 0.26)	
SDQ total difficulties (17)	23.68 (4.5)	17.28 (7.4)	23.70 (4.6)	13.71 (4.9)	3.238** (0.91 to 5.57)	0.57
SDQ impact	0.93 (1.4)	0.57 (1.2)	0.69 (1.1)	0.16 (0.6)	0.288 (-0.08 to 0.66)	
ECBI-I (127)	146.49 (38.9)	136.17 (43.5)	138.04 (28.1)	113.75 (31.9)	20.272** (4.95 to 35.59)	0.55
ECBI-P (11)	17.94 (9.2)	14.70 (9.9)	15.95 (8.9)	10.20 (8.8)	2.748 (-0.50 to 5.99)	

* indicates significant at $p < .05$

** indicates significant at $p < .01$